DOI:
10.1038/nature19792

Document Version
Peer reviewed version

The transferability of lipid loci across African, Asian and European cohorts

Karoline Kuchenbaecker[1,2,3]*, Nikita Telkar[4], Theresa Reiker[3,5,6,7], Robin G. Walters[8,9], Kuang Lin[9], Anders Eriksson[10], Deepti Gurdasani[3], Arthur Gilly[3,11], Lorraine Southam[3,11,12], Emmanouil Tsafantakis[13], Maria Karaleftheri[14], Janet Seeley[15,16,17], Anatoli Kamali[17], Gershim Asiki[17,18,19], Iona Y. Millwood[8,9], Michael Holmes[8,9], Huaidong Du[8,9], Yu Guo[20], Understanding Society Scientific Group[+], Meena Kumari[21], George Dedoussis[22], Liming Li[23], Zhengming Chen[9], Manjinder S. Sandhu[24], Eleftheria Zeggini[3,11]

[1] Division of Psychiatry, University College of London, London W1T 7NF, UK

[2] UCL Genetics Institute, University College London, London WC1E 6BT, UK

[3] Department of Human Genetics, Wellcome Sanger Institute, Hinxton, CB10 1SA, UK

[4] Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

[5] Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK

[6] Swiss Tropical and Public Health Institute, Basel, Switzerland

[7] University of Basel, Basel, Switzerland

[8] Medical Research Council Population Health Research Unit (MRC PHRU), Nuffield Department of Population Health, University of Oxford, Oxford, UK

[9] Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK

[10] Department of Medical and Molecular Genetics, King's College London, London SE1 9RT, UK

[11] Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[12] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

[13] Anogia Medical Centre, Anogia, 740 51, Greece

[14] Echinos Medical Centre, Echinos, Xanthi 67300, Greece

[15] Department of Global Health and Development, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

[16] Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK

[17] Medical Research[1] Council/Uganda Virus Research Institute and London School of Hygiene (MRC/UVRI and LSHTM), Uganda Research Unit, Entebbe, Uganda

[18] African Population and Health Research Center, Nairobi, Kenya

[19] Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

[20] Chinese Academy of Medical Sciences, Beijing 100730, China

[21] Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, UK

[22] Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Greece

[23] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China

[24] Department of Medicine, University of Cambridge, Cambridge CB2 0QQ, UK

[+] A full list of consortium members appears at the end of this manuscript

* Correspondence and requests for materials should be addressed to K. K. (email: k.kuchenbaecker@ucl.ac.uk)

Abstract

Most genome-wide association studies are based on samples of European descent. We assess whether the genetic determinants of blood lipids, a major cardiovascular risk factor, are shared across populations. Genetic correlations for lipids between European-ancestry and Asian cohorts are not significantly different from 1. A genetic risk score based on LDL-cholesterol-associated loci has consistent effects on serum levels in samples from the UK, Uganda and Greece ($r=0.23$-0.28, $p<1.9\times10^{-14}$). Overall, there is evidence of reproducibility for ~75% of the major lipid loci from European discovery studies, except triglyceride loci in the Ugandan samples (10% of loci). Individual transferable loci are identified using trans-ethnic colocalization. Ten of fourteen loci not transferable to the Ugandan population have pleiotropic associations with BMI in Europeans; none of the transferable loci do. The non-transferable loci might affect lipids by modifying food intake in environments rich in certain nutrients, which suggests an important role for gene-environment interactions.

Introduction

Genome-wide association studies (GWAS) have been very successful in identifying genetic variants linked to cardiovascular disease (CVD) and to cardiometabolic traits[1]. Due to the improving predictive accuracy of these variants, genetic risk prediction could soon be implemented in clinical settings[2,3]. However, the majority of samples included in these genome "white" association studies were British or US-Americans with European ancestry[4,5] which does not accurately represent the ethnically and ancestrally diverse populations of these nations. Moreover, three quarters of CVD-associated deaths occur in low- and middle-income countries where incidences are rising[6]. Consequently, it is important to determine whether cardiometabolic loci are transferable to other populations.

Previous research assessed the effects of different allele frequencies and linkage disequilibrium (LD) on genetic associations across ancestry groups[7]. Here we ask the fundamental question whether causal variants for blood lipids, a major cardiovascular risk factor, are shared across populations. Heterogeneity in effects of variants could result from epistasis or gene-environment interactions. However, the causal variants are usually unknown. The differences in LD structure between populations make it difficult to compare the observed associations between ancestry groups because the effect of a variant depends on its correlation with the causal variant(s)[7]. Differences in allele frequency also impact the power to detect associations in other ancestry groups.

We employ several strategies which account for these effects and do not require knowledge of the specific causal variants to quantify the extent to which genetic variants affecting lipid biomarkers are shared between individuals from Europe/North America, Asia, and Africa. We assess the transferability of individual signals and compare association patterns across the genome using data from the African Partnership for Chronic Disease Research – Uganda (APCDR-Uganda, N=6,407)[8], China Kadoorie Biobank (CKB, N=21,295)[9], the Hellenic Isolated Cohorts (HELIC-MANOLIS, N=1,641 and HELIC-Pomak, N=1,945)[10,11], and the UK Household Longitudinal Study (UKHLS, N=9,961)[12]. We also use summary statistics from Biobank Japan (BBJ, N=162,255)[13] and the Global Lipid Genetics Consortium (European ancestry, GLGC2013 N=188,577, GLGC2017 N=237,050)[14,15]. We find evidence

for extensive sharing of genetic variants that affect levels of HDL- and LDL-cholesterol and triglycerides between individuals with European ancestry and samples from China, Japan and Greek population isolates. We estimate that about three quarters of major lipid loci are reproducible. Using trans-ethnic colocalization, we show that many established loci for triglycerides do not affect levels of this biomarker in Ugandan samples, however. Ten out of fourteen of the lipid loci that were not transferable to the Ugandan samples had pleiotropic associations with BMI in European ancestry samples. None of the transferable loci were linked to BMI. This could suggest an important role of environmental factors in modifying which genetic variants affect lipid levels.

Results

Reproducibility of established lipid loci

We assessed rates at which established lipid-associated variants were reproducible in other populations. We selected major lipid loci, i.e. those with lipid associations at $p<10^{-100}$ based on a score test in the largest European ancestry GWAS. In this context, reproducibility was operationalised as at least one variant from the credible set being associated at $p<10^{-3}$ based on a score test with serum lipid levels in the target study. We defined the credible set as variants correlated at $r^2>0.6$ with the lead SNP from the European discovery study. Correlation was estimated from the 1000 Genomes Project samples with European ancestry. As a benchmark, we also assessed replication in a European ancestry study, UKHLS. We found evidence of transferability for 76.5% of major HDL loci in this study (Table 1). For the non-European groups rates ranged from 70.6 to 82.4%. Similar reproducibility rates were observed for LDL loci (61.5-76.9%). For major triglycerides (TG) loci, rates ranged from 78.9 to 94.7%, except in APCDR-Uganda. Only 10.5% of the TG loci showed evidence of reproducibility in that sample. Rates for known loci with $p\geq10^{-100}$ in the discovery set were generally below 10%. However, Biobank Japan, the largest study, exhibited markedly higher reproducibility rates for these loci than the other studies with 24.6-32.7%.

Trans-ethnic genetic correlations

Trans-ethnic genetic correlations were estimated between the three largest studies, China Kadoorie Biobank, Biobank Japan and GLGC2013 (Figure 1). For GLGC2013 and BBJ, correlations were 0.999, 0.778, 0.999 for HDL, LDL and TG, respectively. For GLGC2013 and CKB, correlations were 0.999, 0.959, 0.961 for HDL, LDL and TG, respectively. None of the estimates were significantly different from 1 (Supplementary Table 1). We also compared associations across lipid biomarkers. This consistently showed negative genetic correlations between TG associations and HDL associations, with estimates ranging from r_{gen}=-0.48 to r_{gen}=-0.86.

Genetic risk scores

In order to assess patterns of sharing of risk alleles for the smaller studies, we constructed genetic risk scores (GRS) based on the established lipid loci from discovery studies with European-ancestry and validated the score associations with serum levels of HDL, LDL and TG in HELIC, APCDR-Uganda, CKB and also UKHLS as a benchmark (Figure 2). All genetic scores were significantly associated with their respective target lipid in the three European samples with largely consistent correlation coefficients and mutually overlapping 95% confidence intervals (CIs) (Table 2). For HDL, LDL and TG, the estimated correlation coefficients ranged from 0.27-0.28, 0.23-0.28 and 0.20-0.24, respectively. In APCDR-Uganda, the strongest association was observed for LDL (r=0.28, SE=0.01, $p=1.9x10^{-107}$ based on a mixed model score test). The HDL association was attenuated compared to the European ancestry samples (r=0.12, SE=0.01, p=6.1x10-^{22}). The effect of the TG score was markedly weaker (r=0.06, SE=0.01, $p=4.5x10^{-7}$). For CKB, the HDL GRS had a correlation of r=0.18 (SE=0.02, $p=1.4x10^{-22}$) and the LDL GRS of r=0.20 (SE=0.02, $p=32x10^{-26}$) while the triglyceride GRS showed a stronger attenuation relative to UKHLS with r=0.14 (SE=0.02, $p=3.8x10^{-12}$). We also assessed associations between a given score and levels of each of the other lipid biomarkers (Supplementary Table 2). In line with the trans-

ethnic genetic correlation results, we observed inverse associations between the HDL score and TG levels and vice versa in all studies, except APCDR-Uganda.

Trans-ethnic colocalization

Differences in LD structure, MAF and sample size make it difficult to assess the transferability of individual loci. Therefore, we propose a new strategy to assess evidence for shared causal variants between two populations: trans-ethnic colocalization. For this we re-purposed a method that was originally developed for colocalization of GWAS and eQTL results: Joint Likelihood Mapping (JLIM)[16]. In order to assess its performance for GWAS results from samples with different ancestry, we carried out a simulation study. UK Biobank (UKB) was used as a reference with European ancestry and compared to CKB and APCDR-Uganda. In order to derive an upper boundary for the power, we compared UKB to the ancestry-matched UKHLS set. Phenotypes were simulated. Effect size estimates were varied between 0.10 and 0.25 in order to represent a range similar to that observed for major lipid loci[15]. In the simulations of distinct causal variants in the non-European and the reference group, the frequencies of false positives were as expected close to 0.05 (Supplementary Table 3, Supplementary Figure 1). The power to detect shared associations for betas of 0.25 was 73.1% for APCDR-Uganda, 93.1% for CKB and 0.89 for UKHLS (Figure 3). To investigate whether the lower power for APCDR-Uganda could be due to its smaller sample size, we reran the analyses for CKB using a random subset of samples matching the sample size of APCDR-Uganda. For effect sizes less than 0.2, the results from this analysis revealed decreased detection power relative to the full CKB set but still consistently higher than APCDR-Uganda. This suggests that the power of this trans-ethnic colocalization method decreases somewhat with greater genetic distance between the populations that are compared.

We applied trans-ethnic colocalization for established lipid loci to each study with UKHLS as the reference. There was evidence for significant (p_{jlim}<0.05 based on a permutation test) colocalization with at least one of the target studies for about half of the major lipid loci (Supplementary Table 4).

For several of the major TG loci, such as 8q24.13, strong evidence of transferability to the Asian studies was observed whilst there was no evidence of association in APCDR-Uganda. Figure 4 shows the regional association plots of this locus for each data set as an example to demonstrate that differences in LD and frequencies lead to different association patterns. As colocalization can account for such differences, the result from the analysis comparing the European and Asian studies was nevertheless statistically significant ($p < 0.001$).

We compared major lipid loci that showed evidence of transferability to APCDR-Uganda with those that did not. The proximal genes of transferable loci were enriched for lipid pathways including lipoprotein metabolism, lipid digestion mobilisation and transport, chylomicron-mediated lipid transport and metabolism of lipids and lipoproteins. The proximal genes of the non-transferable loci were enriched for several other pathways in addition to lipid metabolism, including SHP2 signalling, ABV3 integrin pathway, cytokine signalling in immune system, cytokine-cytokine receptor interaction and transmembrane transport of small molecules (Supplementary Figures 2 and 3). We also assessed the associations of these loci with BMI in samples with European ancestry using publicly available summary statistics from the GIANT consortium[17] ($N \geq 484,680$) (Table 3). Ten of the fourteen non-transferable lipid loci had pleiotropic associations with BMI at a Bonferroni-adjusted threshold of $p<0.0024$. None of the seven transferable lipid loci were associated with BMI.

Discussion

Recent efforts to increase global diversity in genetics studies have been vital, enabling this comprehensive cross-population comparison of genetic associations with blood lipids. We provide evidence for extensive sharing of genetic variants that affect levels of HDL- and LDL-cholesterol and triglycerides between individuals with European ancestry and samples from China, Japan and Greek population isolates. We estimated that about three quarters of major lipid loci are reproducible. This was highly consistent across all studies except for triglyceride loci in APCDR-Uganda. None of the estimates of trans-ethnic genetic correlations between European, Chinese and Japanese samples were

significantly different from 1. All GRS associations in the two Greek isolated populations were highly consistent with those in the UK samples (correlations ranged from 0.27-0.28, 0.23-0.28 and 0.20-0.24, for HDL, LDL and TG, respectively, in these studies). Associations of genetic risk scores for LDL were not attenuated in the Ugandan population compared to the UK samples (r=0.28, SE=0.01, $p=1.9x10^{-107}$ based on a score test).

Previous studies that compared the direction of effect of established loci or assessed associations of genetic risk scores reported differing degrees of consistency[18–29]. However, most of them were conducted in American samples with diverse ancestry, had smaller sample sizes and applied a single-variant look-up or GRS for a limited number of genetic variants. The high degree of consistency for cholesterol biomarkers we observed also contrasts with previously reported trans-ethnic genetic correlations for other traits, such as major depression, rheumatoid arthritis, or type 2 diabetes, which were substantially different from 1[30,31]. In a recent application using data from individuals with European and Asian ancestry from the UK and USA, the average genetic correlation across multiple traits was 0.55 (SE = 0.14) for GERA and 0.54 (SE=0.18) for UK Biobank[32].

As a limitation of our study, we did not adjust for use of lipid-lowering medication. This could in principle cause a small downward bias for the genetic effect estimates. However, few of the participants of the Ugandan and Chinese studies used lipid-lowering drugs. So this is unlikely to have an effect on the main conclusions of this work.

Differences in LD structure, MAF and sample size make it difficult to assess the transferability of individual loci. We therefore propose a new approach: trans-ethnic colocalization. Simulations showed consistent control of type I error rates as well as power greater than 80% to detect shared associations between samples with European and Chinese ancestry for SNP effects greater or equal to 0.15. However, power was decreased for comparisons between samples from APCDR-Uganda and UK Biobank (51.5-73.1%). Hence, for the current implementation non-significant colocalization should not be considered as definitive evidence for the absence of shared causal variants when comparing African and European samples. Future work should address this through better modelling of the LD

structure. Moreover, for many of the major lipid loci, more than one independent association signal was identified in discovery GWASs[15]. When these are located in close proximity to each other, they can interfere with the trans-ethnic colocalization analysis because JLIM assumes a single causal variant. Therefore, future work should extend this approach to accommodate loci harbouring multiple causal variants.

Using trans-ethnic colocalization, we showed that many established loci for triglycerides did not affect levels of this biomarker in Ugandan samples. This included loci associated at genome-wide significance in all the other studies, such as *GCKR* at 2p23.3 or *LPL* at 8p21.3. The genetic risk score for triglycerides had a weak effect on measured levels in APCDR-Uganda. This is unlikely to be an artefact of unreliable measurement: triglyceride levels had a heritable component in this sample (SNP heritability of 0.25, SE=0.05[8]) and there were genome-wide significant associations. It is also unlikely that this can be explained purely by differences in LD and MAF because they would affect the analyses of the other two lipid biomarkers as well. Instead these discrepancies could be caused by gene-environment interactions. Ten out of fourteen of the lipid loci that were not transferable to the Ugandan samples had pleiotropic associations with BMI in European ancestry samples while none of the transferable loci were linked to BMI. It is possible that the non-transferable variants affect the amount of food intake with downstream consequences for lipid levels. This might require an environment offering diets that are rich in certain nutrients. While the proximal genes for transferable loci were almost exclusively linked to pathways of lipid metabolism, the ones for non-transferable loci were involved in diverse pathways which is in line with hypothesis. An alternative explanation could be that the non-transferable loci are involved in metabolising nutrients given a particular diet that is not common in Uganda with downstream consequences for weight.

Overall, this could suggest an important role of environmental factors in modifying which genetic variants affect lipid levels. Studying the causes for discordant loci between groups has promise to further elucidate the biological mechanisms of lipid regulation and other complex traits. Applying genetic risk prediction within clinical settings is receiving increasing attention. Our findings

demonstrate that the transferability of genetic associations across different ancestry groups and environmental settings should be assessed comprehensively for medically relevant traits. This is important in order to ensure that health benefits of precision medicine are widely shared within and across populations. Ongoing programs in underrepresented countries[33], such as the Human Hereditary and Health in Africa Initiative[34], and programs focussing on underrepresented groups, such as PAGE[35], All of Us[36], or East London Genes and Health[37], could provide the basis for this.

Methods

Data resources

We included data from the Global Lipid Genetics Consortium (European ancestry samples only, GLGC), The UK Household Longitudinal Study (UKHLS), two isolated populations from the Greece Hellenic Isolated Cohorts (HELIC) , a rural West Ugandan population from the African Partnership for Chronic Disease Research (APCDR-Uganda) study, China Kadoorie Biobank (CBK), and Biobank Japan (BBJ). Raw genotype and phenotype data were available for UKHLS, APCDR-Uganda, CKB, HELIC-MANOLIS, and HELIC-Pomak. All participants provided written informed consent and each study obtained approval from ethical review boards. The APCDR-Uganda study was approved by the Uganda Virus Research Institute, Science and Ethics Committee (Ref. GC/127/10/10/25), the Uganda National Council for Science and Technology (Ref. HS 870), and the U.K. National Research Ethics Service, Research Ethics Committee (Ref. 11/H0305/5). The HELIC study was approved by the Harokopio University Bioethics Committee. The UKHLS has been approved by the University of Essex Ethics Committee and the nurse data collection by the National Research Ethics Service (10/H0604/2). For CKB, central ethics approvals were obtained from Oxford University, and the China National CDC. In addition, approvals were also obtained from institutional research boards at the local CDCs in the 10 regions. BBJ approved by the ethics committees of RIKEN Center for Integrative Medical Sciences and the Institute of Medical Sciences, the University of Tokyo. Our analyses were based on summary statistics for BBJ and GLGC.

The details of genotyping, QC and imputation for all studies are summarised in Supplementary Table 5. Descriptive information about the sample sets is provided in Supplementary Table 6. Details of the quality control, imputation, genome-wide association analyses and ethical approval have also been previously described for GLGC[14], BBJ[13], HELIC[10], APCDR-Uganda[8] and UKHLS[12]. Each study confirmed sample ethnicity through PCA which rules out sample overlap between studies.

For CKB, 102,783 participants were genotyped using 2 custom-designed Affymetrix Axiom® arrays including up to 803K variants, optimized for genome-wide coverage in Chinese populations. Stringent quality control included SNP call rate >0.98, plate effect $P>10^{-6}$, batch effect $P>10^{-6}$, HWE $P>10^{-6}$ (combined 10df χ^2 test from 10 regions), biallelic, MAF difference from 1KGP EAS < 0.2, sample call rate >0.95, heterozygosity <mean+3SD, no chrXY aneuploidy, genetically-determined sex concordant with database, resulting in genotypes for 532,415 variants present on both array versions. Imputation into the 1,000 Genomes Phase 3 reference (EAS MAF>0) using SHAPEIT version 3 and IMPUTE version 4 yielded genotypes for 10,276,633 variants with MAF >0.005 and info >0.3.

In CKB, lipid levels were regressed against eight principle components, region, age, age^2, sex, and - for LDL and TG - fasting time^2 for the single SNP association analysis. For CKB, PCs were included in both single SNP and PRS association analyses to improve inflation. Recruitment for CKB occurred at 10 different rural and urban locations across China leading to somewhat increased population structure. The resulting inflation estimates lambda after PC adjustment were 1.063, 1.050, and 1.053 for HDL, LDL and TG, respectively. LDL levels were derived using the Friedewald formula. After rank-based inverse normal transformation, the residuals were used as the outcomes in the genetic association analyses using linear regression. Associations were carried out within a mixed model framework using BOLT-LMM[38].

The single SNP association analysis for APCDR-Uganda was carried out within a mixed model framework using GEMMA[39]. Rank-based inverse normal transformation was applied to the lipid biomarkers after adjusting for age and gender. For Uganda, the inflation estimates lambda were 1.000, 1.004, and 1.005 for HDL, LDL and TG, respectively.

Established lipid loci

A list of established lipid-associated loci was extracted from the latest Global Lipid Genetics Consortium (GLGC2017) publication[15] reporting 444 independent variants in 250 loci associated at genome-wide significance with HDL, LDL, and triglyceride levels. We excluded three LDL variants where the association was not primarily driven by the samples with European ancestry. We assessed evidence for transferability of the loci, applied trans-ethnic colocalization and used them to construct genetic risk scores.

Reproducibility of established lipid loci

We assessed evidence that these established lipid signals generalise to other populations. For loci harbouring multiple signals, we only kept the most strongly associated variant. Out of the 444 loci, this left 170 HDL, 135 LDL and 136 TG variants. We distinguished major loci, i.e. those with $p<10^{-100}$ based on a score test in GLGC2017. For each lead SNP we identified all variants in LD ($r^2>0.6$) based on the European ancestry 1000 Genomes data. We assessed whether the lead or any of the correlated variants, henceforth called credible set, displayed evidence of association in the target study. If this was not the case, we tested whether there was any other variant with evidence of association within a 50Kb window. We used a p-value threshold of $p<10^{-3}$ based on a score test. This threshold was derived by computing the minimum p-value in 1000 random windows of 50Kb for each study. Less than 5% of random windows had a minimum $p<10^{-3}$ for the non-European ancestry studies. While this p-value threshold might not be appropriate to provide conclusive evidence of reproducibility for individual loci, we used this to test evidence of reproducibility across sets of loci. These analyses excluded the HELIC studies because the smaller sample size makes it difficult to differentiate between lack of power and lack of reproducibility.

Trans-ethnic genetic correlations

We used the popcorn software[30] to estimate trans-ethnic genetic correlations between studies while accounting for differences in LD structure. This provides an indication of the correlation of causal-variant effect sizes across the genome at SNPs common in both populations. Variant LD scores were estimated for ancestry-matched 1000 Genomes v3 data for each study combination. The estimation of LD scores failed for chromosome 6 for some groups. We therefore left out the major histocompatibility complex (MHC) region (positions 28,477,797 to 33,448,354) from chromosome 6 from all comparisons. Variants with imputation accuracy r^2<0.8 or MAF<0.01 were excluded. Popcorn did not converge for any of the studies with less than 20,000 samples. Therefore, results are presented for comparisons between GLGC2013, CKB and BBJ. We estimated effect rather than impact correlations. We used a Bonferroni correction to adjust for multiple testing of three traits with each other ($p<0.05/9=0.0056$).

Genetic risk scores

As it was not possible to compute trans-ethnic genetic correlations for UKHLS, the HELIC cohorts, and APCDR-Uganda, we created genetic risk scores based on the established lipid loci and assessed their associations with serum lipid levels in these studies. We also tested the associations of GRS in CKB as raw data were available for this study as well. Age and sex were adjusted for by regressing them on the lipid biomarker values and using the residuals as outcomes for subsequent analyses. For CKB, we additionally adjusted for 20 PCs and region covariates in order to ensure population structure was accounted for. To ensure values are normally distributed, we used rank-based inverse normal transformation for all biomarkers and data sets which involves ordering values first and then assigning them to expected normal values. To make sure GRS were comparable across studies, we excluded variants that were absent, rare (MAF<0.01) or badly imputed (r^2<0.8) in any of the studies and variants that had different alleles from those in the GLGC. The variant with larger discovery p-value from each correlated pair of SNPs (r^2>0.1) was also removed. These filters were applied based on each, UKHLS, HELIC, and APCDR-Uganda and then the intersection of variants was carried forward to generate GRS.

Out of the 444 loci, this left 120, 103 and 101 variants for HDL, LDL and TG, respectively (Supplementary Table 7). We created trait-specific weighted GRS. The β-regression coefficients from SNP-trait associations in GLGC2017[15] were used as weights. All lipid biomarkers and scores were scaled to mean=0 and standard deviation=1 for each study, so that the regression coefficients represent estimates of the correlation between scores and lipid biomarkers.

We carried out association analyses between each genetic risk score and each lipid biomarkers using a linear mixed model with random polygenic effect implemented in GEMMA[39] in order to account for relatedness and population structure. For CKB, we used BOLT-LMM because it is efficient for large samples. We used a Bonferroni correction to adjust for multiple testing of three GRS with three different lipid biomarker outcomes (p<0.05/9=0.0056 for the score test).

Trans-ethnic colocalization

Differences in allele frequency, LD structure and sample size make it difficult to assess whether a given GWAS hit are transferable to samples with different ancestries. Therefore, we applied trans-ethnic colocalization. Colocalization methods test whether the associations in two studies can be explained by the same underlying signal even if the specific causal variant is unknown. The joint likelihood mapping (JLIM) statistic was developed by Chun and colleagues to estimate the posterior probabilities for colocalization between GWAS and eQTL signals and compare them to probabilities of distinct causal variants[16]:

$$\Lambda = \sum_{i \in N_\theta^1(m^*)} L_1(i) \times \log \frac{L_1(i)L_2(i)}{\max\limits_{j \notin N_\theta^2(i)} L_1(i)L_2(j)} \qquad (1)$$

i SNP
m^* lead SNP
$L_1(i)$ likelihood of SNP i being causal for trait 1
$L_2(i)$ likelihood of SNP i being causal for trait 2
$N_\theta^1(i)$, $N_\theta^2(i)$ sets of SNPs in LD with *i*
θ LD threshold

JLIM explicitly accounts for LD structure. Therefore, we assessed whether it is suitable for trans-ethnic colocalization. For the reference sample set, it was possible to use genome-wide summary statistics

for the analysis. For this set, LD scores were estimated using a subset of samples from the 1000 Genomes Project v3 that had matching ancestry to that study. The second sample set needed raw genotype data and LD was estimated directly for these samples. JLIM assumes only one causal variant within a region in each study. We therefore used small windows of 50Kb for each known locus to minimise the risk of interference from additional association signals. Distinct causal variants were defined by separation in LD space by $r^2 \geq 0.8$ from each other. We excluded loci within the MHC region due to its complex LD structure. We used a significance threshold of $p<0.05$ given the evidence of association of the established lipid loci in Europeans and the overall evidence for shared causal genetic architecture across populations for most lipid traits from our other analyses. We compared each target study to UKHLS because of the study's high level of homogeneity in terms of ancestry, biomarker quantification and study design.

Simulation

To test the power of trans-ethnic colocalization to detect associations shared between pairs of populations with different ancestry, we ran JLIM on two sets of simulated traits with realistic effect size and environmental noise level. The first set of simulations used the same causal variant in both populations, whereas the second set of simulations used discordant causal variants. Causal variants were selected using the sample function in R, corresponding to a uniform random draw from the entire chromosome. We sampled 10,000 randomly chosen biallelic variants with MAF>0.05 and simulated random phenotypes in UKHLS, CKB, APCDR-Uganda and 50,000 individuals with British ancestry from UK Biobank as the reference set. For UK Biobank we applied the QC and used the ancestry assignment provided by Bycroft et al[40]. UKHLS was included as an ancestry-matched set in order to derive an upper limit estimate of the power. For each data set relatives were excluded. We also sub-sampled CKB to match the number of individuals in APCDR-Uganda in order to test whether the difference in performance was due to ancestry or sample size. We used a simple linear model to generate the phenotype for each individual *i*:

(2)

$$y_i = \beta * (x_i - 1) + \eta_i$$

where y is the phenotype value, β is the effect size, x is the number of the alternate alleles carried at the locus and $\eta_i \sim N(0, \sigma^2)$, where σ^2 is the variance of the environmental noise and $\mathrm{Cov}(\eta_i, \eta_j) = 0$. We tested effect size estimate beta from 0.10, 0.15, 0.20 and 0.25 in order to represent a range similar to that observed for the major lipid loci[15]. We used σ^2 = 1 to match the trait variances of the standardised phenotypes.

Comparison of transferable loci with non-transferable loci

We assessed whether there are any systematic differences between loci that are shared between European ancestry samples and APCDR-Uganda and loci that are not. We identified all loci with evidence of reproducibility based on the above definition that also had significant ($p<0.05$) colocalization based on a permutation test. We only kept one variant per region. We contrasted them with loci where none of the evidence suggested generalisation: $p>0.05$ for colocalization or missing result due to failed convergence, no variant with a lipid association at $p<10^{-3}$ in the region and the lead variant from the discovery study was not rare in APCDR-Uganda. We identified the nearest protein coding gene for each locus and carried out pathway analyses for the two sets using FUMA[41]. We also assessed the associations of the lead variants with body mass index (BMI) in European ancestry samples using results from a meta-analysis between the GIANT consortium and UK Biobank[17]. We used a Bonferroni adjusted p-value threshold.

Data availability

The UKHLS EGA accession number is EGAD00010000918. Genotype-phenotype data access for UKHLS is available by application to Metadac (www.metadac.ac.uk). Summary statistics for GLGC (http://csg.sph.umich.edu/abecasis/public/) and Biobank Japan (http://jenger.riken.jp/en/) are publicly available. The HELIC genotype and WGS datasets have been deposited to the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/home): EGAD00010000518;

EGAD00010000522; EGAD00010000610; EGAD00001001636, EGAD00001001637. The APCDR committees are responsible for curation, storage, and sharing of the APCDR-Uganda data under managed access. The array and sequence data have been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/, study accession number EGAS00001000545, datasets EGAS00001001558 and EGAD00001001639 respectively) and can be requested through datasharing@sanger.ac.uk. Requests for access to phenotype data and summary statistics may be directed to data@apcdr.org. This is restricted to research-related purposes. Uploading and sharing of individual genetic data from CKB are subject to restrictions according to the Interim Measures for the Administration of Human Genetic Resources administered by the Human Genetic Resources Administration of China (HGRAC). Summary data including allele frequencies and GWAS summary statistics are available on application. This is restricted to research-related purposes. Other individual-level CKB data are available through www.ckbiobank.org, subject to completion of a Material Transfer Agreement, either through Open Access or on application. CKB data access is subject to oversight by an independent Data Access Committee.

Code availability

Our code to run trans-ethnic colocalization using JLIM and simulations is available through github: https://github.com/KarolineKuchenbaecker/TEColoc

Acknowledgements

CKB thanks the participants, project staff, the China National Centre for Disease Control and Prevention and its regional offices. The Chinese National Health Insurance scheme provided electronic linkage to all hospital admission data. We thank the residents of the Pomak villages and of the Mylopotamos villages for taking part. We thank the African Partnership for Chronic Disease Research (APCDR) for providing a network to support this study as well as a repository for deposition of curated data. We also thank all study participants who contributed to this study. UKHLS is led by the Institute

for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The survey was conducted by NatCen and the genome-wide scan data were analysed and deposited by the Wellcome Trust Sanger Institute. This work was funded by the Wellcome Trust (WT098051), (212360/Z/18/Z), and the European Research Council (ERC-2011-StG 280559-SEPI). The baseline survey and first resurvey for CKB were supported by a research grant from the Hong Kong Kadoorie Charitable Foundation. Long-term follow-up and the second resurvey were supported by grants from the UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), National Natural Science Foundation of China (81390540, 81390541, 81390544), and National Key Research and Development Program of China (2016YFC 0900500, 0900501, 0900504, 1303904). DNA extraction and genotyping was supported by grants from GlaxoSmithKline and the UK Medical Research Council (MC-PC-13049, MC-PC-14135). MVH is supported by the British Heart Foundation (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. The British Heart Foundation, UK Medical Research Council, and Cancer Research UK provide core funding to the Clinical Trial Service Unit and Epidemiological Studies Unit, Oxford University (Oxford, UK). APCDR-Uganda was funded by the Wellcome Trust, The Wellcome Trust Sanger Institute (WT098051), the UK Medical Research Council (G0901213-92157, G0801566, and MR/K013491/1), and the Medical Research Council/Uganda Virus Research Institute Uganda Research Unit on AIDS core funding. The UK Household Longitudinal Study was funded by grants from the Economic & Social Research Council (ES/H029745/1) and the Wellcome Trust (WT098051).

Author contributions

KK conceived this project and supervised the work. KK and NT carried out the genetic correlation and PRS analyses. KK carried out all analyses involving trans-ethnic colocalization. KK wrote the manuscript. KK and AE carried out the simulation study. TR implemented earlier versions of the genetic risk scores. HELIC: EZ and GD are the principle investigators, AG and LS carried out the quality control, MK and ET were involved in data collection. APCDR-Uganda: MS, DG, GA, JS, AK were involved

in collecting and preparing data as well as leading the study. China Kadoorie Biobank: ZC and LL are the principle investigators; RGW is the genomics lead; RGW, IYM, HD, YG and MVH were involved in data collection; RGW and KL carried out quality control and genome-wide association analysis for lipid biomarkers; KL carried out the genotype imputation. All authors approved the manuscript.

Competing interests

The Authors declare no competing interests.

References

1. Clarke, S. L. & Assimes, T. L. Genome-Wide Association Studies of Coronary Artery Disease: Recent Progress and Challenges Ahead. *Curr. Atheroscler. Rep.* **20**, (2018).
2. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
3. Kuchenbaecker, K. B. *et al.* Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J. Natl. Cancer Inst.* **109**, (2017).
4. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
5. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nat. News* **538**, 161 (2016).
6. Bovet, P. & Paccaud, F. Cardiovascular Disease and the Changing Face of Global Public Health: A Focus on Low and Middle Income Countries. *Public Health Rev.* **33**, 397–415 (2011).
7. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
8. Heckerman, D. *et al.* Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7377–7382 (2016).
9. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* **40**, 1652–1666 (2011).

10. Gilly, A. *et al.* Cohort-wide deep whole genome sequencing and the allelic architecture of complex traits. *Nat. Commun.* **9**, 4674 (2018).
11. Gilly, A. *et al.* Very low depth whole genome sequencing in complex trait association studies. *Bioinforma. Oxf. Engl.* (2018). doi:10.1093/bioinformatics/bty1032
12. Prins, B. P. *et al.* Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Sci. Rep.* **7**, 11008 (2017).
13. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
14. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
15. Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758 (2017).
16. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600–605 (2017).
17. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
18. Musunuru, K. *et al.* Multi-Ethnic Analysis of Lipid-Associated Loci: The NHLBI CARe Project. *PLoS ONE* **7**, (2012).
19. Teslovich, T. M. *et al.* Biological, Clinical, and Population Relevance of 95 Loci for Blood Lipids. *Nature* **466**, 707–713 (2010).
20. Dumitrescu, L. *et al.* Genetic Determinants of Lipid Traits in Diverse Populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *PLoS Genet.* **7**, (2011).
21. Bryant, E. K. *et al.* A Multiethnic Replication Study of Plasma Lipoprotein Levels-Associated SNPs Identified in Recent GWAS. *PLoS ONE* **8**, (2013).
22. Wu, Y. *et al.* Trans-Ethnic Fine-Mapping of Lipid Loci Identifies Population-Specific Signals and Allelic Heterogeneity That Increases the Trait Variance Explained. *PLoS Genet.* **9**, (2013).

23. Wang, Z. *et al.* Genetic associations with lipoprotein subfraction measures differ by ethnicity in the multi-ethnic study of atherosclerosis (MESA). *Hum. Genet.* **136**, 715–726 (2017).
24. Keebler, M. E. *et al.* Association of Blood Lipids with Common DNA Sequence Variants at Nineteen Genetic Loci in the Multiethnic United States National Health and Nutrition Examination Survey III. *Circ. Cardiovasc. Genet.* **2**, 238–243 (2009).
25. Lanktree, M. B., Anand, S. S., Yusuf, S. & Hegele, R. A. Replication of genetic associations with plasma lipoprotein traits in a multiethnic sample. *J. Lipid Res.* **50**, 1487–1496 (2009).
26. Chang, M. *et al.* Racial/ethnic variation in the association of lipid-related genetic variants with blood lipids in the US adult population. *Circ. Cardiovasc. Genet.* **4**, 523–533 (2011).
27. Below, J. E. *et al.* Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs. *Sci. Rep.* **6**, (2016).
28. Johnson, L., Zhu, J., Scott, E. R. & Wineinger, N. E. An Examination of the Relationship between Lipid Levels and Associated Genetic Markers across Racial/Ethnic Populations in the Multi-Ethnic Study of Atherosclerosis. *PloS One* **10**, e0126361 (2015).
29. Marigorta, U. M. & Navarro, A. High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants. *PLoS Genet.* **9**, e1003566 (2013).
30. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).
31. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
32. Galinsky, K. J. *et al.* Estimating cross-population genetic correlations of causal effect sizes. *Genet. Epidemiol.* (2018). doi:10.1002/gepi.22173
33. Hindorff, L. A. *et al.* Prioritizing diversity in human genomics research. *Nat. Rev. Genet.* **19**, 175–185 (2018).
34. Mulder, N. *et al.* H3Africa: current perspectives. *Pharmacogenomics Pers. Med.* **Volume 11**, 59–66 (2018).

35. Carlson, C. S. *et al.* Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. *PLOS Biol.* **11**, e1001661 (2013).
36. Program Overview - All of Us | National Institutes of Health. Available at: https://allofus.nih.gov/about/about-all-us-research-program. (Accessed: 28th December 2018)
37. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community based population genomics and health study in people of British-Bangladeshi and -Pakistani heritage. Preprint at https://doi.org/10.1101/426163 (2018).
38. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).
39. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
40. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
41. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

The Understanding Society Scientific Group are Michaela Benzeval(25), Jonathan Burton(25), Nicholas Buck(25), Annette Jäckle(25), Heather Laurie(25), Peter Lynn(25), Stephen Pudney(25), Birgitta Rabe(25), Dieter Wolke(26)

25) Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester CO4 3SQ UK , UK

26) Department of Psychology, University of Warwick, Coventry CV4 7AL, UK

Figure legends

Figure 1. Trans-ethnic genetic correlations r_{gen} for associations with high-density lipoprotein (HDL), low-density lipoprotein (LDL) cholesterol and triglycerides (TG). a) shows the comparison of GLGC2013 (European) and Biobank Japan (BBJ), b) GLGC2013 and China Kadoorie Biobank (CKB) and c) BBJ and CKB. The values on the diagonal show correlations for matched lipid markers in the two studies. The off-diagonal values show genetic correlations across lipid biomarkers (e.g. HDL vs TG). Colours correspond to the direction and strength of r_{gen}.

Figure 2. Associations of genetic risk scores based on established lipid-associated loci with serum levels of high-density lipoprotein (HDL), low-density lipoprotein (LDL) cholesterol and triglycerides (TG) in a) UKHLS, b) HELIC-MANOLIS, c) HELIC-Pomak, d) APCDR-Uganda, e) CKB. Estimates are given as correlation coefficients r with colours corresponding to the direction and strength of r. Stars indicate statistically significant associations based on a score test ($p<0.0056$). The values on the diagonal represent the strength of correlation of a GRS with its target lipid biomarker. The off-diagonal elements show the strength of correlation of a GRS (e.g. TG) with the other lipid markers (e.g. HDL).

Figure 3. Power of trans-ethnic colocalization to detect shared associations for different effect sizes (Beta). Based on a simulation study comparing 50,000 samples with European ancestry from UK Biobank to UKHLS, APCDR-Uganda ("UG"), and China Kadoorie Biobank ("CKB"). Additionally, China Kadoorie Biobank was downsampled to N=4597 ("CKB_4K") to match the sample size of APCDR-Uganda.

Figure 4. Regional plots for SNP associations with triglyceride levels at established triglyceride-associated locus 8q24.13 for UKHLS, Biobank Japan (BBJ), APCDR-Uganda, and China Kadoorie Biobank (CKB) and p-value p_{jlim} based on a permutation test for the trans-ethnic colocalization with UKHLS. Filling colour of the points corresponds to the strength of linkage disequilibrium (r2) of each variant with the lead variant rs2954029.

Tables

Table 1. Percentage of established lipid-associated loci with evidence of reproducibility in target studies. Only one SNP was kept for each locus with multiple associated variants in close proximity. Regions were defined as 25Kb either side of the lead variant. The credible set contains the reported lead variant and variants in LD (r2>0.6) with it. Results are shown separately for groups of loci by strength of association (whether $p<10^{-100}$) in the discovery study (GLGC). There were 25, 16 and 27 loci with $p<10^{-100}$ based on a score test in GLGC for HDL, LDL, and TG, respectively. There were 212, 171 and 158 loci with $p\geq10^{-100}$ for HDL, LDL, and TG, respectively.

P in GLGC:		**$<10^{-100}$**			**$\geq10^{-100}$**		
Study	**Trait**	**n.s.***	**Region†**	**Credible‡**	**n.s.***	**Region†**	**Credible‡**
UKHLS	HDL	5.9	17.6	76.5	81.0	13.7	5.2
	LDL	7.7	15.4	76.9	77.0	16.4	6.6
	TG	0.0	5.3	94.7	82.1	14.5	3.4
CKB	HDL	11.8	5.9	82.4	71.2	16.4	12.4
	LDL	7.7	30.8	61.5	83.6	7.4	9.0
	TG	5.3	15.8	78.9	82.9	10.3	6.8
BBJ	HDL	11.8	11.8	76.5	47.7	19.6	32.7
	LDL	7.7	30.8	61.5	64.8	10.7	24.6
	TG	5.3	10.5	84.2	55.6	12.8	31.6
APCDR-Uganda	HDL	11.8	17.6	70.6	73.2	25.5	1.3
	LDL	23.1	7.7	69.2	73.8	24.6	1.6
	TG	42.1	47.4	10.5	79.5	17.1	3.4

* No variant in the region associated in target set at $p<10^{-3}$

† No variant in the credible set associated in the target set at $p<10^{-3}$ but an uncorrelated variant in the region is associated in target set at $p<10^{-3}$

‡ A variant in the credible set is associated in the target set at $p<10^{-3}$

Table 2: Associations of genetic risk scores based on established lipid-associated loci and respective serum lipid levels in UKHLS, HELIC-MANOLIS, -Pomak, APCDR-Uganda, and CKB using a linear mixed model analysis. P-values are based on score tests.

Trait	N	Correlation (SE*)	Confidence interval	P-value
UKHLS				
HDL	9706	0.285 (0.010)	0.265, 0.305	4.52×10^{-166}
LDL	9767	0.274 (0.010)	0.254, 0.294	1.32×10^{-155}
Triglycerides	9635	0.204 (0.010)	0.183, 0.223	9.62×10^{-87}
HELIC-MANOLIS				
HDL	1186	0.279 (0.029)	0.222, 0.336	4.08×10^{-20}
LDL	1186	0.229 (0.029)	0.172, 0.286	2.41×10^{-14}
Triglycerides	1176	0.235 (0.030)	0.176, 0.294	4.52×10^{-14}
HELIC-Pomak				
HDL	1078	0.268 (0.030)	0.209, 0.327	2.39×10^{-17}
LDL	1075	0.290 (0.030)	0.231, 0.349	3.04×10^{-19}
Triglycerides	1066	0.234 (0.030)	0.175, 0.293	2.00×10^{-13}
APCDR-Uganda				
HDL	6407	0.121 (0.012)	0.098, 0.145	6.06×10^{-22}
LDL	6407	0.280 (0.012)	0.257, 0.304	1.91×10^{-107}
Triglycerides	6407	0.063 (0.013)	0.038, 0.089	4.46×10^{-7}
CKB				
HDL	20810	0.180 (0.018)	0.145, 0.215	1.4×10^{-22}
LDL	17662	0.198 (0.019)	0.161, 0.235	3.2×10^{-26}
Triglycerides	20222	0.139 (0.020)	0.100, 0.178	3.8×10^{-12}

*SE=standard error

Table 3. Association of established lipid-associated loci with body mass index by whether the locus was transferable to APCDR-Uganda. BMI association results are based on N≥484,680 samples from the meta-analysis between GIANT and UK Biobank[17]. Results are shown exclusively for loci where there was clear evidence for or against transferability to APCDR-Uganda (see Methods for more details).

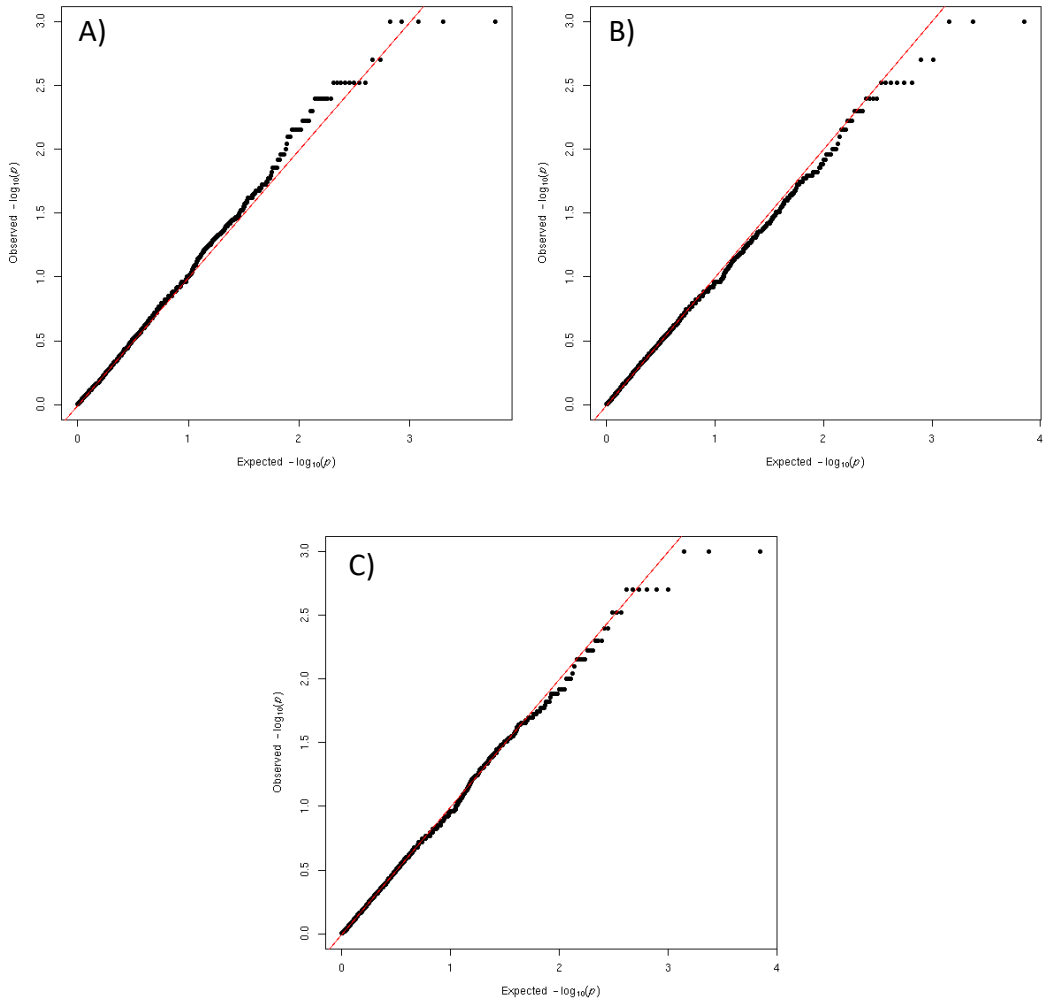
Transferable	Trait	rs-id	Chr	Position	Annotation	beta	SE	P-value
no	HDL	rs11755393	6	34824636	*UHRF1BP1*	-0.025	0.002	**$9.8x10^{-48}$**
no	HDL	rs1178979	7	72856430	*BAZ1B*	-0.010	0.002	**$3.1x10^{-6}$**
no	HDL	rs4731702	7	130433384	*KLF14*	0.008	0.002	**$3x10^{-7}$**
no	HDL	rs2954033	8	126493746	*NSMCE2*	-0.010	0.002	**$6.4x10^{-8}$**
no	LDL	rs4245791	2	44074431	*ABCG8*	0.002	0.002	0.22
no	LDL	rs3846662	5	74651084	*HMGCR*	0.020	0.002	**$1.9x10^{-35}$**
no	LDL	rs2737229	8	116648565	*TRPS1*	0.014	0.002	**$1.9x10^{-15}$**
no	LDL	rs635634	9	136155000	*IL6R*	0.005	0.002	0.03
no	LDL	rs2000999	16	72108093	*HPR*	0.011	0.002	**$8.6x10^{-8}$**
no	TG	rs1260326	2	27730940	*GCKR*	-0.011	0.002	**$1.2x10^{-10}$**
no	TG	rs2943641	2	227093745	*IRS1*	0.006	0.002	**$5.8x10^{-4}$**
no	TG	rs6905288	6	43758873	*VEGFA*	-0.010	0.002	**$1.9x10^{-9}$**
no	TG	rs11820589	11	116633862	*APOA5*	-0.003	0.003	0.41
no	TG	rs58542926	19	19379549	*TM6SF2*	-0.003	0.003	0.33
yes	HDL	rs643531	9	15296034	*TTC39B*	0.000	0.002	0.92
yes	HDL	rs1800588	15	58723675	*LIPC*	-0.002	0.002	0.25
yes	HDL	rs3764261	16	56993324	*CETP*	-0.002	0.002	0.39
yes	HDL	rs16942887	16	67928042	*PSKH1*	-0.005	0.003	0.06
yes	LDL	rs12740374	1	109817590	*CELSR2*	0.003	0.002	0.18
yes	LDL	rs1367117	2	21263900	*APOB*	-0.002	0.002	0.19
yes	LDL	rs6511720	19	11202306	*LDLR*	0.006	0.003	0.03

"The transferability of lipid loci across African, Asian and European cohorts"

Kuchenbaecker et al.

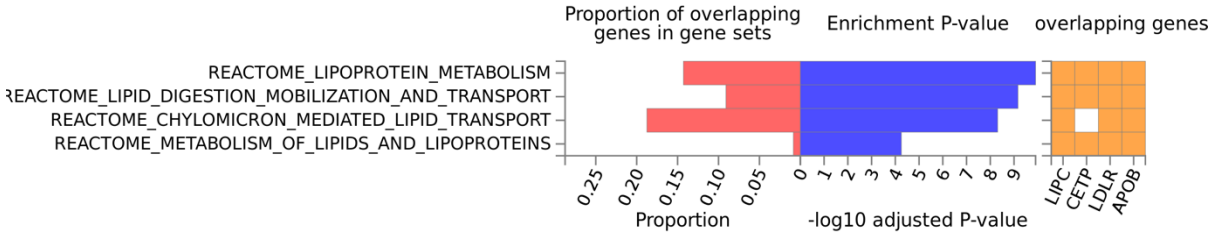
Supplementary Information

Supplementary figures and tables

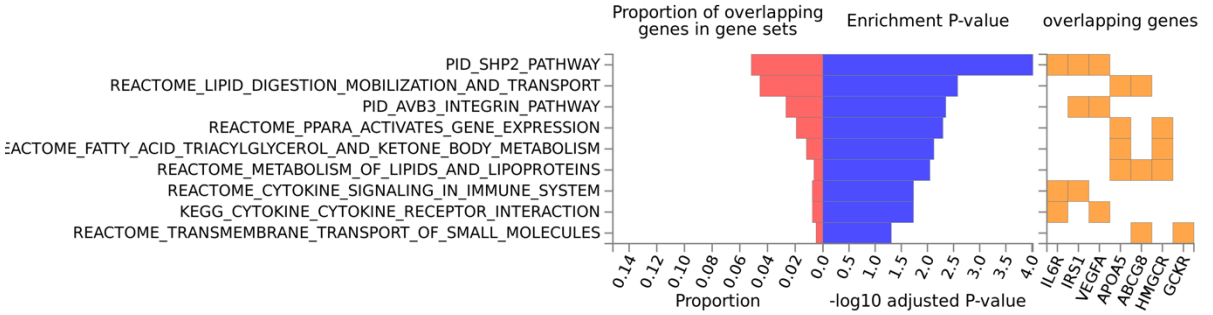


Supplementary Figure 1: QQ-plots for the p-values based on a permutation test from the trans-ethnic colocalization for simulated traits with distinct causal variants in UK Biobank and A) CKB and B) APCDR-Uganda C) UKHLS to provide an ancestry-matched control. Well-controlled type I errors manifest as a close match between the points and the diagonal line.

A)

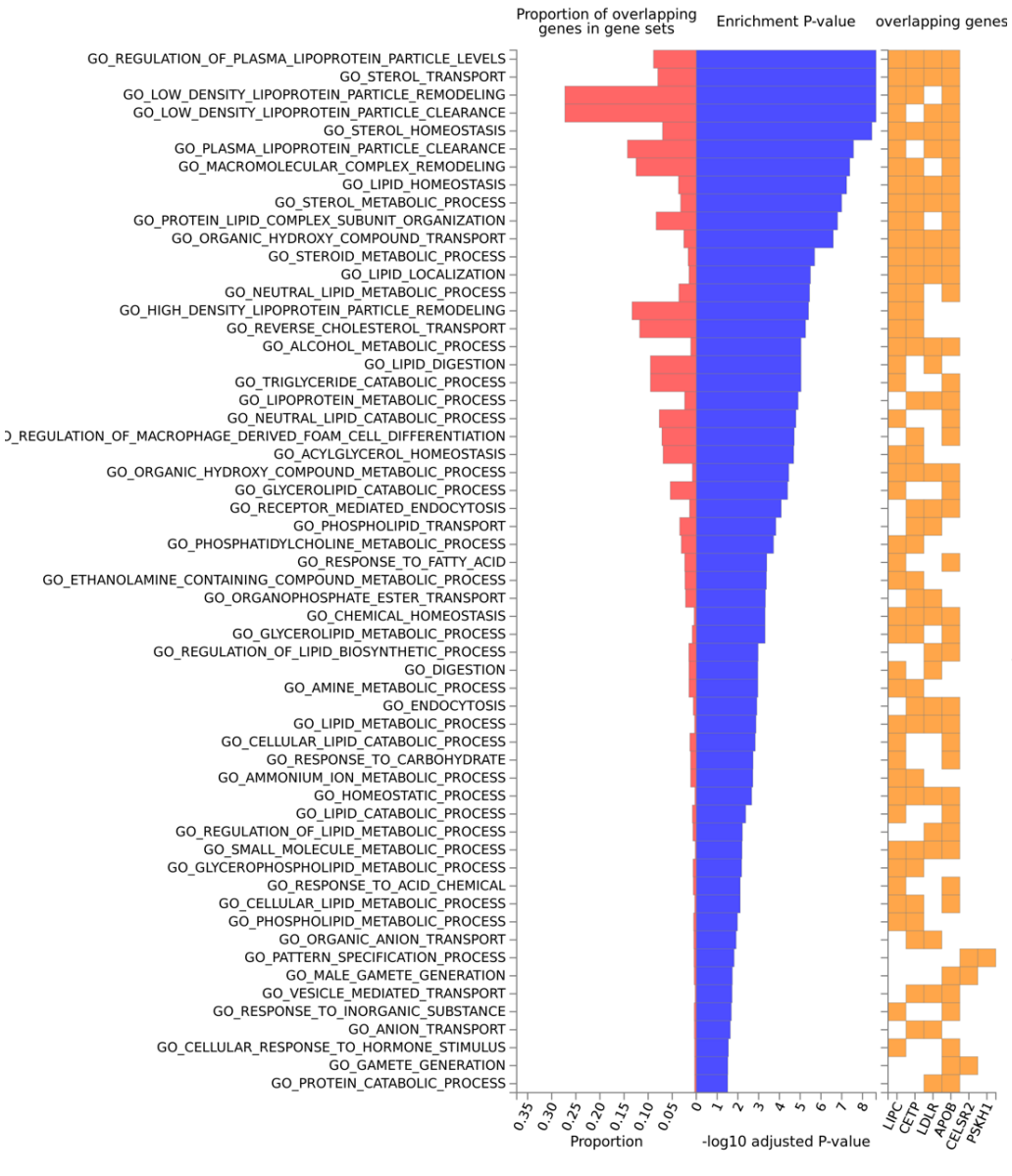


B)



Supplementary Figure 2: Enrichment of canonical pathways (MsigDB c2) for genes proximal to established lipid-associated SNPs that A) replicate in APCDR-Uganda and B) do not replicate in APCDR-Uganda

A)


Proportion of overlapping genes in gene sets
Enrichment P-value
overlapping genes
GO_REGULATION_OF_PLASMA_LIPOPROTEIN_PARTICLE_LEVELS
GO_STEROL_TRANSPORT
GO_LOW_DENSITY_LIPOPROTEIN_PARTICLE_REMODELING
GO_LOW_DENSITY_LIPOPROTEIN_PARTICLE_CLEARANCE
GO_STEROL_HOMEOSTASIS
GO_PLASMA_LIPOPROTEIN_PARTICLE_CLEARANCE
GO_MACROMOLECULAR_COMPLEX_REMODELING
GO_LIPID_HOMEOSTASIS
GO_STEROL_METABOLIC_PROCESS
GO_PROTEIN_LIPID_COMPLEX_SUBUNIT_ORGANIZATION
GO_ORGANIC_HYDROXY_COMPOUND_TRANSPORT
GO_STEROID_METABOLIC_PROCESS
GO_LIPID_LOCALIZATION
GO_NEUTRAL_LIPID_METABOLIC_PROCESS
GO_HIGH_DENSITY_LIPOPROTEIN_PARTICLE_REMODELING
GO_REVERSE_CHOLESTEROL_TRANSPORT
GO_ALCOHOL_METABOLIC_PROCESS
GO_LIPID_DIGESTION
GO_TRIGLYCERIDE_CATABOLIC_PROCESS
GO_LIPOPROTEIN_METABOLIC_PROCESS
GO_NEUTRAL_LIPID_CATABOLIC_PROCESS
O_REGULATION_OF_MACROPHAGE_DERIVED_FOAM_CELL_DIFFERENTIATION
GO_ACYLGLYCEROL_HOMEOSTASIS
GO_ORGANIC_HYDROXY_COMPOUND_METABOLIC_PROCESS
GO_GLYCEROLIPID_CATABOLIC_PROCESS
GO_RECEPTOR_MEDIATED_ENDOCYTOSIS
GO_PHOSPHOLIPID_TRANSPORT
GO_PHOSPHATIDYLCHOLINE_METABOLIC_PROCESS
GO_RESPONSE_TO_FATTY_ACID
GO_ETHANOLAMINE_CONTAINING_COMPOUND_METABOLIC_PROCESS
GO_ORGANOPHOSPHATE_ESTER_TRANSPORT
GO_CHEMICAL_HOMEOSTASIS
GO_GLYCEROLIPID_METABOLIC_PROCESS
GO_REGULATION_OF_LIPID_BIOSYNTHETIC_PROCESS
GO_DIGESTION
GO_AMINE_METABOLIC_PROCESS
GO_ENDOCYTOSIS
GO_LIPID_METABOLIC_PROCESS
GO_CELLULAR_LIPID_CATABOLIC_PROCESS
GO_RESPONSE_TO_CARBOHYDRATE
GO_AMMONIUM_ION_METABOLIC_PROCESS
GO_HOMEOSTATIC_PROCESS
GO_LIPID_CATABOLIC_PROCESS
GO_REGULATION_OF_LIPID_METABOLIC_PROCESS
GO_SMALL_MOLECULE_METABOLIC_PROCESS
GO_GLYCEROPHOSPHOLIPID_METABOLIC_PROCESS
GO_RESPONSE_TO_ACID_CHEMICAL
GO_CELLULAR_LIPID_METABOLIC_PROCESS
GO_PHOSPHOLIPID_METABOLIC_PROCESS
GO_ORGANIC_ANION_TRANSPORT
GO_PATTERN_SPECIFICATION_PROCESS
GO_MALE_GAMETE_GENERATION
GO_VESICLE_MEDIATED_TRANSPORT
GO_RESPONSE_TO_INORGANIC_SUBSTANCE
GO_ANION_TRANSPORT
GO_CELLULAR_RESPONSE_TO_HORMONE_STIMULUS
GO_GAMETE_GENERATION
GO_PROTEIN_CATABOLIC_PROCESS
0.35 0.30 0.25 0.20 0.15 0.10 0.05
0 1 2 3 4 5 6 7 8
Proportion
-log10 adjusted P-value
LIPC
CETP
LDLR
APOB
CELSR2
PSKH1

B)


Proportion of overlapping genes in gene sets
Enrichment P-value
overlapping genes
Proportion
-log10 adjusted P-value

Supplementary Figure 3: Enrichment of GO biological processes (MsigDB c5) for genes proximal to established lipid-associated SNPs that A) replicate in APCDR-Uganda and B) do not replicate in APCDR-Uganda

Supplementary Table 1: Trans-ethnic genetic correlation estimates and p-values based on a Z-test whether the genetic correlation deviates from 1 for each lipid biomarker in one study with each lipid biomarkers in the other study

biomarker	correlation	standard error	p-value*
GLGC2013 (European) – China Kadoorie Biobank			
HDL-HDL	0.999	-	-
LDL-LDL	0.778	0.300	0.460
TG-TG	0.999	-	-
HDL - LDL	0.068	0.229	0.000
HDL - TG	-0.550	0.176	0.000
LDL - HDL	-0.238	0.143	0.000
LDL - TG	0.473	0.155	0.001
TG - HDL	-0.741	0.162	0.000
TG - LDL	-0.226	0.195	0.000
GLGC2013 (European) – Biobank Japan			
HDL-HDL	0.999	0.081	0.999
LDL-LDL	0.959	0.138	0.765
TG-TG	0.961	0.066	0.555
HDL - LDL	-0.055	0.097	0.000
HDL - TG	-0.592	0.130	0.000
LDL - HDL	-0.277	0.150	0.000
LDL - TG	0.294	0.056	0.000
TG - HDL	-0.481	0.176	0.000
TG - LDL	-0.038	0.091	0.000
China Kadoorie Biobank – Biobank Japan			
HDL-HDL	0.999	-	-
LDL-LDL	0.871	0.225	0.566
TG-TG	0.999	-	-
HDL - LDL	0.085	0.141	0.000
HDL - TG	-0.618	0.151	0.000
LDL - HDL	0.290	0.185	0.000
LDL - TG	0.180	0.169	0.000
TG - HDL	-0.862	0.186	0.000
TG - LDL	-0.097	0.153	0.000

* When the estimate is close to the boundary of 1, popcorn cannot compute the standard error and p-value

Supplementary Table 2: Associations of polygenic scores based on established lipid-associated loci and serum levels of each of the other lipid biomarkers in UKHLS, APCDR-Uganda, HELIC-MANOLIS, -Pomak, and CKB using a linear mixed model analysis with a score test to derive the p-value.

score - trait	correlation	SE (β)	p-value
UKHLS			
HDL - LDL	-0.0526	1.01E-02	2.02E-07
HDL - TG	-0.0720	1.02E-02	1.63E-12
LDL - HDL	-0.0516	1.01E-02	3.97E-07
LDL - TG	0.00845	1.02E-02	0.407
TG - HDL	-0.134	1.01E-02	1.67E-39
TG - LDL	0.0117	1.01E-02	0.247
APCDR-Uganda			
HDL - LDL	-0.0149	1.25E-02	0.234
HDL - TG	-0.0144	1.25E-02	0.249
LDL - HDL	-0.0773	1.25E-02	6.75E-10
LDL - TG	0.0422	1.25E-02	7.27E-04
TG - HDL	-0.00513	1.25E-02	0.681
TG - LDL	-0.0467	1.25E-02	1.86E-04
HELIC-Pomak			
HDL - LDL	-0.0722	3.09E-02	1.97E-02
HDL - TG	-0.128	3.10E-02	4.96E-05
LDL - HDL	-0.0608	3.18E-02	6.00E-02
LDL - TG	0.00343	3.19E-02	0.918
TG - HDL	-0.157	3.05E-02	6.16E-07
TG - LDL	0.0739	3.08E-02	1.70E-02
HELIC-Manolis			
HDL - LDL	0.00919	2.96E-02	0.756
HDL - TG	-0.0936	3.01E-02	2.13E-03
LDL - HDL	-0.0229	2.97E-02	0.442
LDL - TG	-0.00136	3.03E-02	0.964
TG - HDL	-0.133	2.98E-02	1.25E-05
TG - LDL	-0.00163	2.98E-02	0.956
CKB			
HDL - LDL	-0.0082	0.0183	0.65
HDL - TG	-0.0667	0.0200	8.6E-04
LDL - HDL	-0.0287	0.0186	0.12
LDL - TG	-0.0144	0.0203	0.65
TG - HDL	-0.0604	0.0183	9.8E-04
TG - LDL	-0.0499	0.0183	6.5E-03

Supplementary Table 3. Type I error rates from a simulation to assess the performance of trans-ethnic colocalization when causal variants are not shared. Phenotypes were simulated 10,000 times for each study and different values of effect sizes. Trans-ethnic colocalization was run to compare each to a reference set of 50,000 samples with British ancestry from UK Biobank. CKB was also down-sampled to match the sample size of APCDR-Uganda.

	Study (N)	**APCDR-Uganda (4,597)**	**UKHLS (9,150)**	**CKB (72,473)**	**CKB (4,597)**
Beta	0.10	0.044	0.043	0.065	0.059
	0.15	0.040	0.057	0.050	0.050
	0.20	0.048	0.059	0.073	0.059
	0.25	0.048	0.050	0.052	0.045

Supplementary Table 4. P-value for the trans-ethnic colocalization based on a permutation test for the JLIM model for established lipid-associated loci in UKHLS, China Kadoorie Biobank (CKB), Biobank Japan (BBJ) and APCDR-Uganda (UG).

				GLGC			reproducible[a]				JLIM p-value[b]		
rs-id	**chr**	**position**	**near gene**	**MAF**	**p-value**	**Multi[c]**	**UKHLS**	**CKB**	**BBJ**	**UG**	**CKB**	**BBJ**	**UG**
HDL													
rs4660293	1	40028180	PABPC4	0.21	6.1E-36	distant	uc	ns	cor	ns	0.8	0.014	NA
rs11755393	6	34824636	UHRF1BP1	0.36	4.2E-23	distant	ns	ns	cor	ns	0.035	0	NA
rs1178979	7	72856430	BAZ1B	0.18	1.3E-26	near	uc	uc	cor	ns	0	0	NA
rs4731702	7	130433384	KLF14	0.46	1.2E-35	distant	ns	cor	cor	ns	0.14	0.003	NA
rs4841132	8	9183596	PPP1R3B	0.9	1.0E-123	no	ns	ns	ns	cor	0.99	0.24	0.16
rs328	8	19819724	LPL	0.098	1.7E-316	near	cor	cor	cor	cor	0.21	0.005	0.62
rs2954033	8	126493746	NSMCE2	0.72	3.0E-61	near	cor	cor	cor	ns	0.94	0.002	NA
rs643531	9	15296034	TTC39B	0.88	3.8E-42	no	ns	ns	uc	uc	NA	NA	NA
rs2066714	9	107586753	ABCA1	0.15	3.6E-31	near	uc	cor	cor	uc	0.97	0.007	0.94
rs1883025	9	107664301	ABCA1	0.26	2.1E-118	near	uc	cor	cor	uc	0.84	0.8	0.97
rs2792751	10	113940329	GPAM	0.73	3.8E-21	near	ns	ns	cor	uc	0.002	0.002	NA
rs7350481	11	116586283	APOA5	0.91	3.2E-100	distant	cor	cor	cor	uc	0	0	0.99
rs964184	11	116648917	ZPR1	0.85	2.6E-217	near	cor	cor	cor	uc	1	1	1
rs10468017	15	58678512	LIPC	0.27	1.8E-306	near	cor	cor	cor	cor	1	0.98	0.99
rs1800588	15	58723675	LIPC	0.24	0	distant	cor	cor	cor	cor	0.007	0.009	0.017
rs247616	16	56989590	CETP	0.31	0	near	cor	cor	cor	cor	0	0	0
rs3764261	16	56993324	CETP	0.31	0	near	cor	cor	cor	cor	0	0	0
rs34065661	16	56995935	CETP	0.005	5.6E-103	near	uc	uc	uc	cor	0	0	0
rs16942887	16	67928042	PSKH1	0.13	9.8E-93	near	ns	ns	cor	cor	0.96	0.29	0.025
rs72836561	17	41926126	CD300LG	0.028	8.1E-111	no	uc	uc	uc	ns	NA	NA	NA
rs7241918	18	47160953	LIPG	0.85	1.2E-104	distant	cor	cor	cor	ns	0.16	1	1
rs116843064	19	8429323	ANGPTL4	0.02	4.8E-146	near	uc	ns	ns	uc	NA	NA	NA
rs769449	19	45410002	APOE	0.11	6.9E-129	near	cor	cor	cor	cor	0.009	0.02	0.95
rs386000	19	54792761	LILRB2	0.22	1.1E-41	distant	cor	uc	cor	cor	0	1	0.71
LDL													
rs11591147	1	55505647	PCSK9	0.015	0.0	near	uc	uc	uc	uc	0.94	0.64	0.92
rs12740374	1	109817590	CELSR2	0.22	0.0	near	cor	cor	cor	cor	0	0	0

rs1367117	2	21263900	APOB	0.28	3.6E-278	near	cor	cor	cor	cor	1	1	0
rs541041	2	21294975	APOB	0.81	1.3E-287	distant	cor	uc	uc	cor	1	1	NA
rs4245791	2	44074431	ABCG8	0.72	1.7E-120	near	uc	ns	ns	ns	NA	0.81	0.99
rs3846662	5	74651084	HMGCR	0.48	3.3E-128	near	cor	cor	cor	ns	0.002	0	1
rs2737229	8	116648565	TRPS1	0.34	8.9E-15	distant	cor	ns	cor	ns	0.015	0.025	NA
rs635634	9	136155000	IL6R	0.19	4.9E-109	near	ns	cor	cor	ns	0.96	0.97	NA
rs2000999	16	72108093	HPR	0.2	4.0E-71	distant	cor	cor	cor	ns	1	0	1
rs6511720	19	11202306	LDLR	0.11	0.0	near	cor	uc	uc	cor	0	NA	0
rs28399654	19	45316588	BCAM	0.027	7.5E-232	distant	cor	uc	uc	cor	1	1	NA
rs7412	19	45412079	APOE	0.075	0.0E+00	near	cor	cor	cor	cor	0	0	0
Triglycerides													
rs10889353	1	63118196	DOCK7	0.33	6.4E-170	no	cor	cor	cor	uc	0	0	0.88
rs676210	2	21231524	APOB	0.26	4.9E-118	near	cor	uc	cor	uc	1	1	1
rs1260326	2	27730940	GCKR	0.63	0.0	near	cor	cor	cor	ns	0	NA	NA
rs2943641	2	227093745	IRS1	0.66	4.9E-33	no	ns	ns	cor	ns	0.006	NA	1
rs6905288	6	43758873	VEGFA	0.59	9.0E-35	near	cor	ns	cor	ns	0	0	NA
rs1178979	7	72856430	BAZ1B	0.18	1.5E-179	near	cor	cor	cor	ns	0	0	NA
rs35332062	7	73012042	MLXIPL	0.12	5.2E-205	distant	cor	cor	cor	uc	0.99	1	NA
rs326	8	19819439	LPL	0.3	0.0	near	cor	cor	cor	uc	0.91	0.91	0.72
rs2954029	8	126490972	TRIB1	0.45	8.3E-205	near	cor	cor	cor	ns	0	0	1
rs1883025	9	107664301	ABCA1	0.26	1.2E-13	no	uc	ns	ns	ns	1	0.001	NA
rs7350481	11	116586283	APOA5	0.91	0.0	distant	cor	cor	cor	uc	0	0	0
rs11820589	11	116633862	APOA5	0.066	4.4E-133	near	cor	uc	uc	ns	0	1	1
rs2075291	11	116661392	APOA5	0.003	5.7E-65	near	uc	cor	cor	ns	NA	1	NA
rs10047462	11	116722041	SIK3	0.86	9.9E-180	near	cor	cor	cor	uc	0	0	1
rs247616	16	56989590	CETP	0.31	2.4E-38	near	ns	ns	cor	cor	0.014	0.024	0.78
rs116843064	19	8429323	ANGPTL4	0.02	4.2E-175	near	uc	ns	ns	ns	NA	NA	NA
rs58542926	19	19379549	TM6SF2	0.074	3.7E-125	no	cor	cor	cor	ns	NA	NA	NA
rs439401	19	45414451	APOE	0.63	2.7E-168	near	cor	cor	cor	uc	0.009	0.53	1

[a] indicates whether any variant from the credible set ("cor") or any uncorrelated variant within 50kb ("uc") is associated with the target biomarker at $p<10^{-3}$ in each of the target studies

[b] p-value from the JLIM trans-ethnic colocalization analysis using UKHLS as the comparison set

[c] indicates whether multiple independent hits have been reported within 50kb ("near") or 1Mb ("distant")

Supplementary Table 5: Summary of the genotyping, quality control, and imputation of each study

study	array	QC criteria SNPs	N SNPs genotyped	N SNPs after QC	Reference panel imputation	Imputation method	N SNPs imputed and QCed	QC criteria samples	N samples after QC
UK Household Longitudinal Study[1]	HumanCoreExome	Hardy-Weinberg equilibrium p-value < 1×10^{-4}, call rate < 98%, poor genotype clustering values (<0.4)	538,448	525,314	UK10K, 1000 Genomes v3	SHAPEIT, IMPUTE2	24,727,032	call rate <98%, autosomal heterozygosity outliers (>3SD), gender mismatches, duplicates (PI_HAT > 0.9), non-European ancestry	9,962 (9798 for HDL, 9797 for LDL, 9807 for TG)
African Partnership for Chronic Disease Research - Uganda[2]	HumanOmni2.5	call rate <0.97, Hardy-Weinberg equilibrium $p<10^{-8}$	2,369,382	2,330,014	1000G v3, 1,978 samples from Uganda	SHAPEIT, IMPUTE2	19,539,450	call rate <97%, heterozygosity (>3SD), gender mismatch, IBD>0.90, ancestry outliers (none)	6,407 (for all biomarkers)
China Kadoorie Biobank[3]	Custom Affymetrix Axiom Array	SNP call rate >0.98, plate effect $P>10^{-6}$, batch effect $P>10^{-6}$, HWE $P>10^{-6}$ (combined 10df χ^2 test from 10 regions), biallelic, MAF difference from 1KGP EAS < 0.2	701K/830K,	532,415	1000 Genomes v3	SHAPEIT3 and IMPUTE4	10,276,633	sample call rate >0.95, heterozygosity <mean+3SD, no chrXY aneuploidy, genetically-determined sex concordant with database	21,295 (20,810 for HDL, 17,662 for LDL, 20,219 for TG)

RIKEN Biobank Japan[4]	HumanOmniExpress	call rate < 0.99, minor allele frequency < 1%, Hardy–Weinberg equilibrium p≤ 1.0×10−6	~1M	NA	1000 Genomes v 3 East Asians	MACH, minimac	6,108,953	call rate <0.98, closely related individuals based on IBD, non–East Asian outliers identified by PCA together with HapMap samples	162,255 (70,657 for HDL, 72,866 for LDL, 105,597 for TG)
Hellenic Isolated Cohorts[5,6] MANOLIS, Pomak	Whole-genome sequencing	VQSR with a tranche threshold of 99.4%, call rate <99%	NA	NA	NA	NA	24,163,896	sex checks, low concordance (π^<0:8) with chip data, duplicates, traces of contamination; checked but no exclusion necessary: depth, heterozygosity, transition/transversion (Ti/Tv) rate, missingness, ethnicity.	1,641 (1632 for HDL, 1630 for LDL, 1632 for TG), 1,945 (1915 for HDL, 1914 for LDL, 1916 for TG),
Global Lipids Genetics Consortium [7]	23 studies GWAS arrays, 37 Metabochip	study specific	196,710	study specific	HapMap	MACH	2.6M	study specific	188,577
Global Lipids Genetics	HumanExome		NA	242,289	NA	NA	NA	call rate, heterozygosity, sex discordance, GWAS discordance,	237,050

Consortium [8]								fingerprint concordance, PCA outliers	

Supplementary Table 6: Study description and mean levels of HDL-cholesterol, LDL-cholesterol and triglycerides (TG) in mmol/l

study	acronym	population	array	N SNPs genotyped, imputed	N samples	%female	mean age	mean HDL	mean LDL	mean TG
UK Household Longitudinal Study[1]	**UKHLS**	British	HumanCoreExome	248K, 26M	9,962	44	52	1.55	3.02	1.81
African Partnership for Chronic Disease Research - Uganda[2]	**APCDR-Uganda**	Ugandan	HumanOmni2.5	2.2M, 20M	6,407	56	34	1.02	2.05	1.18
China Kadoorie Biobank[3]	**CKB**	Chinese	Custom Affymetrix Axiom Array	701K/830K, 10M	21,295	62	60	1.37	2.19	1.69
RIKEN Biobank Japan[4]	**BBJ**	Japanese	HumanOmniExpress	6M	162,255	63	43	1.42	3.38	1.50
Hellenic Isolated Cohorts[5,6]	**HELIC-MANOLIS, -Pomak**	Isolated Greek populations	Whole-genome sequencing	24M	1,641, 1,945	42,34	62,45	1.28, 1.18	3.27, 3.09	1.61, 1.58
Global Lipids Genetics Consortium[7]	**GLGC2013** (meta-analysis)	European ancestry	23 studies GWAS arrays, 37 Metabochip	200K, 2.5M	188,577					
Global Lipids Genetics Consortium[8]	**GLGC2017** (meta-analysis)	European ancestry	HumanExome	242,289	237,050					

Supplementary Table 7: SNPs used to create genetic risk scores for HDL-cholesterol, LDL-cholesterol and triglycerides (TG) with reference alleles and weights.

HDL			LDL			TG		
rs-id	reference	weight	rs-id	reference	weight	rsid	reference	weight
rs11553746	C	0.015	rs13379043	T	-0.018	rs1011731	G	-0.015
rs2276853	G	-0.015	rs12748152	C	0.031	rs900399	A	-0.014
rs28932178	T	0.02	rs12740374	G	-0.16	rs10861661	A	0.019
rs740363	G	0.014	rs676210	G	-0.039	rs12748152	C	0.031
rs622082	A	-0.017	rs3756772	C	0.014	rs4846914	G	-0.039
rs2074158	T	-0.02	rs9376090	T	-0.025	rs676210	G	-0.071
rs1011731	G	0.015	rs4841132	A	0.057	rs13389219	C	-0.037
rs900399	A	0.019	rs2293889	T	-0.015	rs2943641	T	0.033
rs9816226	A	0.028	rs2954029	A	-0.048	rs13326165	A	0.02
rs12055786	C	-0.021	rs4149268	C	-0.015	rs9311651	A	-0.021
rs4871137	G	-0.022	rs687621	A	0.043	rs645040	G	0.023
rs10968576	A	-0.017	rs2068888	G	-0.016	rs442177	G	0.031
rs7076938	C	0.019	rs7941030	T	0.014	rs13133548	G	0.014
rs1037378	G	-0.015	rs173539	C	-0.033	rs9686661	C	0.042
rs746463	C	-0.017	rs9939224	T	-0.023	rs459193	A	0.023
rs7136716	A	0.021	rs7241918	G	0.02	rs998584	C	0.034
rs10861661	A	-0.017	rs6511720	G	-0.21	rs2745353	C	0.02
rs10483776	A	-0.02	rs7412	C	-0.54	rs4731702	C	-0.027
rs13379043	T	0.017	rs1132274	C	0.019	rs4841132	A	-0.035
rs8099014	C	0.015	rs4745	A	-0.015	rs1801177	G	0.17
rs2303108	T	-0.015	rs976002	A	0.023	rs2954029	A	-0.08
rs12748152	C	-0.043	rs13146272	C	-0.015	rs2068888	G	-0.032
rs4847399	G	-0.021	rs1016988	T	-0.02	rs2167079	C	-0.02
rs12740374	G	0.045	rs351855	G	-0.018	rs10892063	A	-0.058
rs12145743	T	0.017	rs3812594	G	-0.018	rs1106766	C	-0.03
rs4650994	G	-0.019	rs10885997	A	0.015	rs11057401	T	-0.028
rs1689800	A	-0.025	rs1891110	G	0.021	rs1800588	C	0.047
rs4846914	G	0.049	rs704	G	0.021	rs10468017	C	0.034
rs676210	G	0.06	rs2239619	C	0.018	rs1421085	T	0.019
rs2322659	T	-0.019	rs67710536	A	0.028	rs173539	C	-0.034
rs13389219	C	0.035	rs9646133	G	-0.019	rs9939224	T	-0.034
rs2943641	T	-0.036	rs11080150	A	-0.019	rs2925979	T	-0.029
rs2305637	C	-0.032	rs2125345	T	-0.024	rs2292642	C	-0.02
rs6762477	G	0.025	rs6062343	G	-0.014	rs489693	C	0.015
rs13326165	A	-0.025	rs4809330	A	-0.015	rs891088	A	-0.017
rs9311651	A	0.019	rs10903129	A	0.028	rs7255436	C	-0.019
rs645040	G	-0.021	rs11206510	T	-0.07	rs731839	G	-0.015
rs442177	G	-0.018	rs2479409	G	-0.047	rs7412	C	0.12
rs13133548	G	-0.017	rs505151	G	-0.09	rs7679	T	0.053

rs9686661	C	-0.032	rs10889353	A	-0.045	rs738322	A	-0.02
rs459193	A	-0.02	rs7515577	C	0.03	rs6062343	G	-0.018
rs34525648	G	-0.025	rs267733	A	-0.025	rs10889353	A	-0.077
rs11755393	A	-0.027	rs20558	T	0.015	rs541041	G	0.018
rs2894342	C	0.017	rs2738755	C	-0.015	rs1367117	G	0.023
rs998584	C	-0.026	rs541041	G	0.12	rs4245791	C	-0.019
rs35349911	C	-0.017	rs1367117	G	0.11	rs6882076	T	0.038
rs3756772	C	0.014	rs1801702	C	-0.091	rs1564348	T	0.02
rs2745353	C	-0.023	rs4245791	C	-0.072	rs7758229	G	0.018
rs9376090	T	-0.016	2:44066247	G	-0.11	rs4722551	T	-0.026
rs2303361	T	0.025	rs11556157	A	0.025	rs4921914	C	-0.035
rs4917014	T	0.017	rs2030746	C	0.014	rs2081687	T	-0.019
rs4731702	C	0.033	rs2287623	G	-0.021	rs1935	C	-0.029
rs3735080	C	-0.017	rs887829	C	-0.022	rs2255141	A	0.019
rs4841132	A	0.1	rs2290159	G	-0.021	rs2000999	G	0.021
rs1801177	G	-0.2	rs7640978	C	-0.033	rs11871606	C	0.016
rs2293889	T	0.029	rs2251219	T	0.016	rs58542926	C	-0.12
rs2954029	A	0.035	rs13315871	G	-0.038	rs157580	G	0.047
rs643531	C	0.053	rs3816873	T	-0.017	rs492602	A	0.018
rs4149268	C	-0.034	rs12654264	A	0.066	rs738409	C	-0.018
rs2066714	T	0.043	rs4530754	G	0.017	rs3769823	A	0.017
rs33918808	C	0.071	rs6882076	T	0.039	rs26008	T	-0.028
rs2230808	T	0.027	rs3757354	C	-0.033	rs3803357	C	-0.017
rs687621	A	0.015	rs1264562	G	0.015	rs7946	C	-0.016
rs970548	A	0.026	rs13192471	T	0.038	rs2785990	C	0.016
rs2068888	G	0.023	rs1055569	C	0.019	rs3947	G	0.024
rs2167079	C	0.041	rs1564348	T	0.047	rs3927680	T	-0.018
rs10838738	A	-0.032	rs7770628	C	-0.031	rs7901016	T	0.042
rs499974	C	-0.026	rs7758229	G	0.016	rs797486	C	0.02
rs10892063	A	-0.018	rs12670798	T	0.033	rs7157785	G	0.023
rs7941030	T	0.024	rs4722551	T	0.04	rs1077514	C	0.019
rs7134375	C	0.021	rs4921914	C	-0.022	rs6749689	T	-0.016
rs1106766	C	0.032	rs10102164	G	0.031	rs1049817	A	-0.056
rs7298565	G	0.03	rs2081687	T	-0.028	rs1344642	G	-0.015
rs11057401	T	0.033	rs2737229	A	-0.022	rs3748034	G	0.035
rs838880	C	-0.029	rs11136343	A	0.029	rs6831256	A	0.021
rs1800588	C	0.12	rs3780181	A	-0.037	rs16844401	G	0.03
rs10468017	C	0.11	rs1935	C	0.018	rs1126673	C	0.017
rs34317102	A	0.019	rs2255141	A	-0.028	rs4311394	A	0.018
rs1421085	T	-0.022	rs10128711	T	0.025	rs3873379	T	0.028
rs173539	C	0.23	rs11220462	G	0.043	rs2844480	C	0.023
rs9939224	T	0.2	rs11057830	G	0.023	rs1057373	C	0.03
rs5882	G	-0.092	rs4942486	T	-0.022	rs9271366	G	0.024

rs2925979	T	0.041	rs8017377	G	0.023	rs78957773	C	0.052
rs11869286	G	0.03	rs2000999	G	0.063	rs9472138	C	-0.02
rs2292642	C	0.028	rs34832584	G	0.02	rs4410790	T	0.015
rs7241918	G	0.077	rs314253	T	-0.02	rs2240466	G	-0.12
rs489693	C	-0.019	rs11871606	C	-0.027	rs38855	A	-0.014
rs891088	A	0.015	rs7188	A	0.048	rs11776767	G	0.022
rs2277998	G	0.016	rs11669576	G	0.058	rs326	A	-0.11
rs7255436	C	0.029	rs11557092	T	0.024	rs7940646	T	0.016
rs737337	T	-0.058	rs58542926	C	-0.1	rs174546	C	0.052
rs6511720	G	0.024	rs157580	G	0.072	rs12801636	G	-0.018
rs731839	G	0.017	rs1800437	G	-0.019	rs10047462	G	-0.11
rs2111504	T	0.02	rs492602	A	0.028	rs11820589	G	0.19
rs7412	C	0.098	rs364585	A	0.019	rs3135507	C	0.085
rs17695224	G	-0.028	rs7261862	T	-0.024	rs4149056	T	0.029
rs386000	G	0.054	rs6029526	T	0.035	rs7200543	A	0.024
rs12975366	T	-0.029	rs6016373	A	-0.024	rs12453522	A	0.021
rs1132274	C	-0.02	rs1053593	G	-0.016	rs12947658	A	-0.02
rs6120757	C	-0.017	rs738409	C	-0.018	rs7248104	G	-0.02
rs7679	T	-0.056	rs174546	C	-0.053	rs6818397	T	-0.021
rs181362	C	-0.028	rs2844529	G	0.018			
rs17738527	C	-0.018	rs1169288	A	0.037			
rs4823006	A	0.014						
rs738322	A	0.02						
rs138457	T	0.017						
rs267733	A	0.021						
rs1367117	G	-0.02						
rs2255141	A	-0.027						
rs11871606	C	-0.013						
rs157580	G	-0.026						
rs2785990	C	-0.015						
rs2240466	G	0.043						
rs326	A	0.11						
rs174546	C	-0.042						
rs10047462	G	0.023						
rs11820589	G	-0.087						
rs7200543	A	-0.019						
rs1997243	A	0.026						
rs8060686	T	0.056						

Study references

1. Prins, B. P. *et al.* Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Sci. Rep.* **7**, 11008 (2017).
2. Heckerman, D. *et al.* Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7377–7382 (2016).
3. Chen, Z. *et al.* China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* **40**, 1652–1666 (2011).
4. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).
5. Gilly, A. *et al.* Cohort-wide deep whole genome sequencing and the allelic architecture of complex traits. *Nat. Commun.* **9**, 4674 (2018).
6. Gilly, A. *et al.* Very low depth whole genome sequencing in complex trait association studies. *Bioinforma. Oxf. Engl.* (2018). doi:10.1093/bioinformatics/bty1032
7. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
8. Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758 (2017).